

**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**

Docket No. 11092

Application of:

BHASHYAM, R. et al.

Group Art Unit: 2162

Serial No. 10/661,245

Examiner: CORRIELUS, JEAN M.

Filed: September 12, 2003

For: **CLUSTERING STRINGS**

Commissioner for Patents  
P.O. Box 1450  
Alexandria, VA 22313-1450

**BRIEF ON APPEAL**

Sir:

This is an Appeal Brief in furtherance of the Notice of Appeal filed on January 22, 2009. In light of this Brief, Applicant asks the Board of Patent Appeals and Interferences to reconsider this application.

---

CERTIFICATION OF MAILING UNDER 37 CFR 1.8

I hereby certify that this correspondence is being deposited with the United States Postal Service as first class mail in an envelope addressed to: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450 on \_\_\_\_\_.

By: \_\_\_\_\_  
Name: \_\_\_\_\_

**(I) REAL PARTY IN INTEREST**

The present application is assigned to Teradata US, Incorporated.

**(II) RELATED APPEALS AND INTERFERENCES**

There are currently no known active appeals or interferences related to the present application.

**(III) STATUS OF CLAIMS**

The above-identified patent application was filed on September 12, 2003 with original claims 1 through 12. At present, claims 1-3 and 6-9 are active in the present applications. Claims 1-3 and 6-9 were finally rejected in an Official Action dated August 22, 2008. The final rejection of claims 1-3 and 6-9 is being appealed.

Copies of the claims in their current form are provided in the Claims Appendix (section VIII) of this Appeal Brief.

**(IV) STATUS OF AMENDMENTS**

The U.S. Patent and Trademark Office issued a Final Official Action in the prosecution of the present application on August 22, 2008. In response to this Action, Applicant filed a Notice of Appeal and an Amendment cancelling claims 10-12 on January 22, 2008. An Advisory Action was issued by the U.S. Patent and Trademark Office on February 9, 2009. The Advisory Action indicated that the request for reconsideration had been considered but did not place the application in condition for allowance.

**(V) SUMMARY OF CLAIMED SUBJECT MATTER**Claim 1

Claim 1 recites a method implemented in a computer system, for clustering a string, the string including a plurality of characters, the method including:

identifying R unique n-grams  $T_{1...R}$  in the string (*see, e.g., Application, pg. 3, lines 6-7; Figure 10A, step 1005*);

for every unique n-gram  $T_S$  (*see, e.g., Application, pg. 3, lines 7-8; Figure 10A, step 1010*):

if the frequency of  $T_S$  in a set of n-gram statistics is not greater than a first threshold (*see, e.g., Application, pg. 3, lines 7-8; Figure 10A, step 1015*):

clustering the string with a cluster associated with  $T_S$  (*see, e.g., Application, pg. 3, lines 7-8; Figure 10A, step 1020*);

otherwise:

for every other n-gram  $T_V$  in the string  $T_{1...R, \text{ except } S}$  (*see, e.g., Application, pg. 3, lines 9-12; Figure 10A, step 1025*):

concluding that the frequency of n-gram  $T_V$  is greater than the first threshold (*see, e.g., Application, pg. 3, lines 9-12; Figure 10A, step 1030*), and in response:

if the frequency of n-gram pair  $T_S-T_V$  is not greater than a second threshold (*see, e.g., Application, pg. 3, lines 9-12; Figure 10B, step 1050*):

clustering the string with a cluster associated with the n-gram pair  $T_S-T_V$  (*see, e.g., Application, pg. 3, lines 9-12; Figure 10B, step 1055*);

otherwise:

for every other n-gram  $T_X$  in the string  $T_{1...R}$ , except S and V  
(*see, e.g.*, Application, pg. 3, lines 12-14; Figure  
10B, step 1060):

clustering the string with a cluster associated with  
the n-gram triple  $T_S-T_V-T_X$  (*see, e.g.*,  
Application, pg. 3, lines 12-14; Figure  
10B, step 1065);

where  $T_{1...R}$  is a set of n-grams, R is the number of elements in  
 $T_{1...R}$ , and  $T_S$ ,  $T_V$ , and  $T_X$  are members of  $T_{1...R}$ , and S, V,  
and X are integer indexes to identify members of  $T_{1...R}$ .

#### Claim 6

Claim 6 recites a method implemented in a computer system, for clustering a  
string, the string including a plurality of characters, the method including:

identifying R unique n-grams  $T_{1...R}$  in the string (*see, e.g.*, Application, pg.  
3, lines 26-27; Figure 11A, step 1105);

for every unique n-gram  $T_S$  (*see, e.g.*, Application, pg. 3, lines 27-28;  
Figure 11A, step 1110):

if the frequency of  $T_S$  in a set of n-gram statistics is not greater than a  
first threshold (*see, e.g.*, Application, pg. 3, lines 27-28; Figure  
11A, step 1115):

clustering the string with a cluster associated with  $T_S$  (*see, e.g.*,  
Application, pg. 3, lines 27-28; Figure 11A, step 1120);

otherwise:

for  $i = 1$  to Y (*see, e.g.*, Application, pg. 3, line 29, through pg. 4,  
line 1; Figure 11B, step 1135):

for every unique set of  $i$  n-grams  $T_U$  in the string  $T_{1...R}$ , except  $S$  (*see, e.g., Application, pg. 3, line 29, through pg. 4, line 1; Figure 11B, step 1140*):

if the frequency of the n-gram set  $T_S - T_U$  is not greater than a second threshold (*see, e.g., Application, pg. 3, line 29, through pg. 4, line 1; Figure 11B, step 1145*):

clustering the string with a cluster associated with the n-gram set  $T_S - T_U$  (*see, e.g., Application, pg. 3, line 29, through pg. 4, line 1; Figure 11B, step 1150*);

if the string has not been associated with a cluster with this value of  $T_S$  (*see, e.g., Application, pg. 4, lines 1-4; Figure 11B, step 1125*):

for every unique set of  $Y+1$  n-grams  $T_{UY}$  in the string  $T_{1...R}$ , except  $S$  (*see, e.g., Application, pg. 4, lines 1-4; Figure 11C, step 1165*):

clustering the string with a cluster associated with the  $Y+2$  n-gram group  $T_S - T_{UY}$  (*see, e.g., Application, pg. 4, lines 1-4; Figure 11C, step 1170*),

where  $T_{1...R}$  is a set of n-grams,  $R$  is the number of elements in

$T_{1...R}$ , and  $T_S$ ,  $T_V$ , and  $T_X$  are members of  $T_{1...R}$ , and  $S$ ,  $V$ ,

and  $X$  are integer indexes to identify members of  $T_{1...R}$ .

## **(VI) GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL**

Pursuant to the August 22, 2008 Final Office Action, claims 1-3 and 6-9 stand rejected under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement. The final rejection of claims 1-3 and 6-9 under 35 U.S.C. §112, second paragraph, is being appealed.

The August 22, 2008 Final Office Action and Advisory Action of February 9, 2009 also state that Figures 10A, 10B, 11A, 11B and 11C are not supported by

the original specification. This objection to Figures 10A, 10B, 11A, 11B and 11C is also being appealed.

## **(VII) ARGUMENT**

### Rejection of claims 1-3 and 6-9 under 35 U.S.C. §112, second paragraph

The rejection of claims 1-3 and 6-9 under 35 U.S.C. §112, second paragraph, as failing to comply with the written description requirement is respectfully traversed. The rejection of claims 1-3 and 6-9 under 35 U.S.C. §112, first paragraph, as failing to comply with the written description requirement was first presented in an Office Action dated March 28, 2005. The Final Office Action of August 22, 2008 refers to the Office Action of March 28, 2005 in rejecting claims 1-3 and 6-9 under 35 U.S.C. §112, second paragraph. In the March 28, 2005 Official Action, the Examiner stated:

6. Claims 1-3 and 6-12 are rejected under 35 U.S.C. 112, first paragraph, as failing to comply with the written description requirement. The claim(s) contains subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor(s), at the time the application was filed, had possession of the claimed invention. Claims 1 and 10 recite "identifying R unique n-gram T<sub>1</sub>...R in the string; for every unique n-gram T<sub>S</sub>: if the frequency of T<sub>S</sub> in a set of n-gram statistics is not greater than a first threshold: associating the string with a cluster associated with T<sub>S</sub>; otherwise: for every other n-gram T<sub>V</sub> in the string T<sub>1</sub>...R, except s: if the frequency of n-gram T<sub>V</sub> is greater than the first threshold: if the frequency of n-gram pair T<sub>S</sub>-T<sub>V</sub> is not greater than a second threshold: associating the string with a cluster associated with the n-gram pair T<sub>S</sub>-T<sub>V</sub>; otherwise: for every other n-gram T<sub>X</sub> in the string T<sub>1</sub>...R except s and v: associating the string with a cluster associated with the n-gram triple T<sub>S</sub>-T<sub>V</sub>-T<sub>X</sub>"; and claim 6 recites "identifying R unique n-grams T<sub>1</sub>...R in the string; for every unique n-gram T<sub>S</sub>: if the frequency of T<sub>S</sub> in a set of n-gram

statistics is not greater than a first threshold: associating the string with a cluster associated with  $T_S$ ; otherwise: for  $i = 1$  to  $Y$ : for every unique set of  $i$  n grams  $T_U$  in the string  $T_1...R$ , except  $s$ : if the frequency of the n-gram set  $T_S-T_U$  is not greater than a second threshold: associating the string with a cluster associated with the n-gram set  $T_S-T_U$ ; if the string has not been associated with a cluster with this value of  $T_S$ : for every unique set of  $Y+1$  n-grams  $T_{UY}$  in the string  $T_1...R$ , except  $s$ : associating the string with a cluster associated with the  $Y+2$  n-gram group  $T_S-T_{UY}$ ". The specification page 6, line 19 through pages 8, line 15 as indicated by the Applicants does not provide any detail of the above-mentioned limitations of the claim.

In response to the March 28, 2005 Office Action, Applicants filed an amendment and remarks on July 28, 2005. Now, as in the remarks of July 28, 2005, Applicants assert that the Examiner did not established that he has "a reasonable basis for questioning the adequacy of the disclosure to enable a person of ordinary skill in the art to make and use the claimed invention without resorting to *undue experimentation*." MPEP § 2106.01 (emphasis in original). In particular, the Examiner has not presented "a factual analysis of [the] disclosure to show that a person skilled in the art would not be able to make and use the claimed invention without resorting to undue experimentation." MPEP § 2106.02.

In any case, Applicants believe that the claimed subject matter is generally described in the specification as originally filed on page 6, line 19 through page 8, line 15, which includes pseudocode (page 7, line 23 through page 8, line 12), and in Figs. 4, 5 and 6. An example is provided on page 8, lines 24-32 and Figs. 7, 8 and 9. An example SQL implementation flow for the string clustering technique is provided in Table 1 on page 9. Claims 1-3 and 6-9 are enabled because a person of ordinary skill in the art would know how to make and/or use the claimed invention based on these passages and the other material in the specification. The

pseudo-code on pages 7 and 8 uses different variable names from those used in the claims. Applicants do not know of any requirement and the Examiner has not cited any requirement that the specification use identical variable names as the claims. Applicants are similarly unaware of any requirement that Applicants must use identical variable names in the figures and the claims. In any case, paragraphs [0004] - [0010] of the specification, and Figures 10A, 10B, 11A, 11B and 11C added in an amendment filed on July 28, 2005, include variables with identical names to those used in the claims. The Examiner has not shown how the difference in variable naming between some portions of the specification and the claims renders the specification unable to "enable a person of ordinary skill in the art to make and use the claimed invention without resorting to *undue experimentation*." MPEP § 2106.01 (emphasis in original).

Although Applicants remarks presented in the reply filed on July 28, 2005 have not been specifically acknowledged or addressed by the Examiner, the rejection of claims 1-3 and 6-9 under 35 U.S.C. §112, second paragraph, as failing to comply with the written description requirement did not appear in any of the ten subsequent Office Actions and Advisory Actions dated between October 20, 2005 and March 3, 2008, and claims 1-3 and 6-9 were indicated as allowable in the Office Action dated March 3, 2008.

It is believed that the rejection of claims 1-3 and 6-9 under 35 U.S.C. §112, second paragraph, as failing to comply with the written description requirement is improper, and that claims 1-3 and 6-9 are in condition for allowance.



Objection to Figures 10A, 10B, 11A, 11B and 11C

The objection to Figures 10A, 10B, 11A, 11B and 11C as not supported by the original specification is respectfully traversed.

An objection to the drawings as not showing every feature of the invention specified in the claims was presented in the Office Action of March 28, 2005. The drawings were amended to include Figures 10A, 10B, 11A, 11B and 11C in the response to this Office Action filed on July 28, 2005.

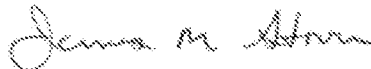
Applicant respectfully traverses the statement in the Final Official Action of August 22, 2008 and the Advisory Action dated February 9, 2009 that the drawings filed on July 28, 2005 (Figures 10A, 10B, 11A, 11B, and 11C) are not supported by the original specification. Each one of Figures 10A, 10B, 11A, 11B, and 11C is supported by disclosure in the application as originally filed. Specifically, Figures 10A and 10B are disclosed, sans reference numerals, in claim 1 as originally filed and in paragraph [0004]; and Figures 11A, 11B, and 11C are disclosed, without reference numerals, in claim 6 as originally filed and in paragraphs [0008]-[0009]. Additional support for Figures 10A, 10B, 11A, 11B, and 11C is provided in Figures 4, 5, 6A, and 6B; paragraphs [0025]-[0034] of the specification; pseudo-code on page 7, line 23-page 8, line 12; and example SQL statements in Table 1 on page 9.

The support for Figures 10A, 10B, 11A, 11B, and 11C was presented in responses filed by Applicants on July 28, 2005 in response to the March 28, 2005 Office Action, and on January 22, 2009 in response to the August 22, 2008 Final Office Action.

Applicants have filed a dozen Office Action replies, including numerous amendments and arguments, to address the persnickety rejections presented in the numerous Office Actions and Advisory Actions received in the prosecution of the present application. Applicants believe that all issues raised by the Examiner have

been addressed and that the application including claims 1-3 and 6-9, is in condition for allowance. Early and favorable action is respectfully requested.

Respectfully submitted,

A handwritten signature in cursive script, appearing to read "James M. Stover", positioned above a horizontal line.

James M. Stover

Reg. No. 32,759

Teradata Corporation  
2835 Miami Village Drive  
Miamisburg, Ohio 45342  
Tel. No. (937) 242-4727

**(VIII) CLAIMS APPENDIX**

1. (previously presented) A method implemented in a computer system, for clustering a string, the string including a plurality of characters, the method including:

identifying  $R$  unique  $n$ -grams  $T_{1...R}$  in the string;

for every unique  $n$ -gram  $T_S$ :

if the frequency of  $T_S$  in a set of  $n$ -gram statistics is not greater than a first threshold:

clustering the string with a cluster associated with  $T_S$ ;

otherwise:

for every other  $n$ -gram  $T_V$  in the string  $T_{1...R}$ , except  $S$ :

concluding that the frequency of  $n$ -gram  $T_V$  is greater than the first threshold, and in response:

if the frequency of  $n$ -gram pair  $T_S$ - $T_V$  is not greater than a second threshold:

clustering the string with a cluster associated with the  $n$ -gram pair  $T_S$ - $T_V$ ;

otherwise:

for every other  $n$ -gram  $T_X$  in the string  $T_{1...R}$ , except  $S$  and  $V$ :

clustering the string with a cluster associated with the  $n$ -gram triple  $T_S$ - $T_V$ - $T_X$ ;

where  $T_{1...R}$  is a set of  $n$ -grams,  $R$  is the number of elements in

$T_{1...R}$ , and  $T_S$ ,  $T_V$ , and  $T_X$  are members of  $T_{1...R}$ , and  $S$ ,  $V$ ,

and  $X$  are integer indexes to identify members of  $T_{1...R}$ .

2. (original) The method of claim 1 further including compiling  $n$ -gram statistics.

3. (original) The method of claim 1 further including compiling n-gram pair statistics.

4. (canceled)

5. (canceled)

6. (previously presented) A method implemented in a computer system, for clustering a string, the string including a plurality of characters, the method including:

identifying  $R$  unique n-grams  $T_{1...R}$  in the string;

for every unique n-gram  $T_S$ :

if the frequency of  $T_S$  in a set of n-gram statistics is not greater than a first threshold:

clustering the string with a cluster associated with  $T_S$ ;

otherwise:

for  $i = 1$  to  $Y$ :

for every unique set of  $i$  n-grams  $T_U$  in the string  $T_{1...R, \text{ except } S}$ :

if the frequency of the n-gram set  $T_S-T_U$  is not greater than a second threshold:

clustering the string with a cluster associated with the n-gram set  $T_S-T_U$ ;

if the string has not been associated with a cluster with this value of  $T_S$ :

for every unique set of  $Y+1$  n-grams  $T_{UY}$  in the string  $T_{1...R, \text{ except } S}$ :

clustering the string with a cluster associated with the  $Y+2$  n-gram group  $T_S-T_{UY}$ ,

where  $T_{1...R}$  is a set of n-grams,  $R$  is the number of elements in

$T_{1...R}$ , and  $T_S$ ,  $T_V$ , and  $T_X$  are members of  $T_{1...R}$ , and  $S$ ,  $V$ , and  $X$  are integer indexes to identify members of  $T_{1...R}$ .

7. (original) The method of claim 6 where  $Y = 1$ .
8. (original) The method of claim 6 further including compiling n-gram statistics.
9. (original) The method of claim 6 further including compiling n-gram group statistics.
10. (canceled)
11. (canceled)
12. (canceled)

**(IX) EVIDENCE APPENDIX**

Not applicable

**(X) RELATED PROCEEDINGS APPENDIX**

Not applicable